

除了前面讨论的 local, flat, vlan 这几类网络，OpenStack 还支持 vxlan 和 gre 这两种 overlay network。

overlay network 是指建立在其他网络上的网络。 该网络中的节点可以看作通过虚拟（或逻辑）链路连接起来的。   
overlay network 在底层可能由若干物理链路组成，但对于节点，不需要关心这些底层实现。

例如 P2P 网络就是 overlay network，隧道也是。 vxlan 和 gre 都是基于隧道技术实现的，它们也都是 overlay network。

目前 linux bridge 只支持 vxlan，不支持 gre；open vswitch 两者都支持。 vxlan 与 gre 实现非常类似，而且 vxlan 用得较多，所以本教程只介绍 vxlan。

VXLAN 全称 Virtual eXtensible Local Area Network。

正如名字所描述的，VXLAN 提供与 VLAN 相同的以太网二层服务，但是拥有更强的扩展性和灵活性。与 VLAN 相比，VXLAN 有下面几个优势：

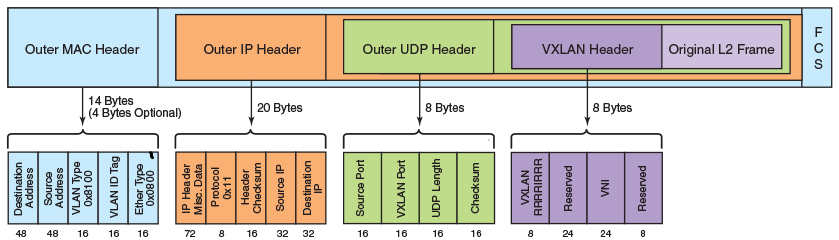
1. 支持更多的二层网段。 VLAN 使用 12-bit 标记 VLAN ID，最多支持 4094 个 VLAN，这对于大型云部署会成为瓶颈。  
   VXLAN 的 ID （VNI 或者 VNID）则用 24-bit 标记，支持 16777216 个二层网段。
2. 能更好地利用已有的网络路径。 VLAN 使用 Spanning Tree Protocol 避免环路，这会导致有一半的网络路径被 block 掉。  
   VXLAN 的数据包是封装到 UDP 通过三层传输和转发的，可以使用所有的路径。
3. 避免物理交换机 MAC 表耗尽。 由于采用隧道机制，TOR (Top on Rack) 交换机无需在 MAC 表中记录虚拟机的信息。

**VXLAN 封装和包格式**

VXLAN 是将二层建立在三层上的网络。 通过将二层数据封装到 UDP 的方式来扩展数据中心的二层网段数量。   
VXLAN 是一种在现有物理网络设施中支持大规模多租户网络环境的解决方案。 VXLAN 的传输协议是 IP + UDP。

VXLAN 定义了一个 MAC-in-UDP 的封装格式。 在原始的 Layer 2 网络包前加上 VXLAN header，然后放到 UDP 和 IP 包中。   
通过 MAC-in-UDP 封装，VXLAN 能够在 Layer 3 网络上建立起了一条 Layer 2 的隧道。

VXLAN 包的格式如下：



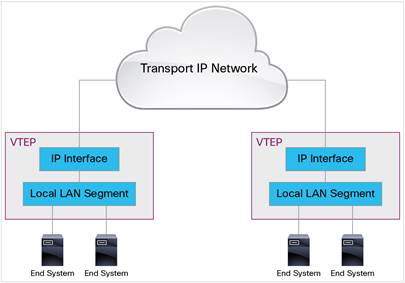
如上图所示，VXLAN 引入了 8-byte VXLAN header，其中 VNI 占 24-bit。 VXLAN 和原始的 L2 frame 被封装到 UDP 包中。

这 24-bit 的 VNI 用于标示不同的二层网段，能够支持 16777216 个 LAN。

**VXLAN Tunnel Endpoint**

VXLAN 使用 VXLAN tunnel endpoint (VTEP) 设备处理 VXLAN 的封装和解封。 每个 VTEP 有一个 IP interface，配置了一个 IP 地址。  
VTEP 使用该 IP 封装 Layer 2 frame，并通过该 IP interface 传输和接收封装后的 VXLAN 数据包。

下面是 VTEP 的示意图：



VXLAN 独立于底层的网络拓扑； 反过来，两个 VTEP 之间的底层 IP 网络也独立于 VXLAN。   
VXLAN 数据包是根据外层的 IP header 路由的，该 header 将两端的 VTEP IP 作为源和目标 IP。

下节我们通过例子进一步理解 VXLAN 封装和转发包的过程，并探讨 Linux 对 VXLAN 的支持。